EXPRESS MAIL LABEL NO.: EV268061051US DATE OF DEPOSIT: DECEMBER 11, 2003

I hereby certify that this paper and fee are being deposited with the United States Postal Service Express Mail Post Office to Addressee service under 37 CFR § 1.10 on the date indicated below and is addressed to the Mail Stop Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

ROBERTA SHERMAN
NAME OF PERSON MAILING PAPER AND FEE SIGNATURE OF PERSON MAILING PAPER AND FEE

Inventor(s): REZA GHASEMI
WALTER HAENEL

A QUALITY EVALUATION TOOL FOR DYNAMIC VOICE PORTALS

BACKGROUND OF THE INVENTION

Statement of the Technical Field

[0001] The present invention relates to the field of computer software and speech recognition and more particularly to user-navigated dynamic voice portals that use speech recognition technology.

Description of the Related Art

[0002] Contrary to visual applications, voice-based applications have the problem that for input recognition no strict pattern matching can be used. The nature of speech recognition makes it very difficult to distinguish between terms having similar pronunciations. Therefore, during the design of speech applications, care should be taken to provide input choices which are pronounced as differently as possible, so as to avoid the problem of recognizing the wrong choice.

[0003] The problem of recognizing the wrong input choice in a speech recognition application occurs with voice portals, which are generally built by various parties that

may not be aware of the terms used in the various applications disposed within the voice portal. Often, a voice portal will have, in addition to the current grammars (or commands) for the actual choice to be made, additional active grammars, such as certain "universal" grammars that allow a user to navigate through the portal, e.g. a command such as "go back." Thus, at any given moment, a combined set of grammars are active, and the voice recognition engine has to search in the set of combined active grammars for a match.

[0004] A problem arises if the various grammars used across the various applications on the portal are designed by different parties, as is the case for voice portals built on a general portal architecture, such as the IBM WebSphere™ Portal Server. General portal architecture allows for new applications to be added dynamically by an administrator. The new added choices created by each new application modify the available choices in a selection menu, and thereby affect the quality of recognition. Generally, the administrators are not voice technology specialists, and may further have to operate a voice portal in multiple languages. Because of this, there is always a risk that a new voice application may drastically reduce the quality of the portal.

[0005] FIG. 1 depicts an example of a sample content and organization of a voice portal. The user is generally presented with a tree 10, into which, after logging into the portal, the user starts at a home directory 11. The tree then divides into new sub-directories 12 and 14, for "Business" and "Entertainment", respectively. At home directory 11, the user would be presented with two choices, for "Business" or "Entertainment," which would be the current grammars for the choice that the portal would need to recognize. In addition to those current grammars, there may be

additional active grammars, such as "go back" or "quit." As the user navigates deeper into the menu 10, the current grammars may change from one menu selection step to another. After the "Places" menu selection step 60, the user would proceed to the "Pages" step 65, and would be presented with a new set of menu options 16, 17, 81, and 19, labeled "Information," Notes, "Directory," and "Sports," respectively. The new menu options would be added to the set of active grammars.

[0006] Below these menu options are the various portlets or voice applications in the applications phase 70 at the bottom of the menu. Applications 20, 22, 24 each branch off from menu item 16, while applications 40, 42, and 44 each branch off from menu item 18. The two sets of voice applications may have been written and arranged by different parties not knowing which terms the other party used for the title of each application. Within each branch of applications additional grammars would be added to the active set which the speech recognition engine of the portal must recognize.

[0007] In menu 10, it can be seen that application 34 is titled "Directory," which is the same as menu option 18. If the grammar for selecting menu option 18 is active within the selection choice following menu option 17, then the system would have trouble distinguishing between identically pronounced terms. Similarly, if a universal grammar such as "store settings" was also active, this would present recognition problems if the user were to navigate through menu item 18, which has the application named "Stores."

[0008] Currently, the only way of testing a portal's recognition quality after setting up the portal or installing a new voice application (or portlet) is to call into the system and check manually, or by user testing with a human user, how well the system works. This

can be time-consuming and expensive. It would be desirable therefore, to provide a quality evaluation tool that assesses the ability of a voice portal to recognize different terms in the various applications attached to the portal, by analyzing and measuring the similarity of the terms.

SUMMARY OF THE INVENTION

[0009] The present invention addresses the deficiencies of the art with respect to evaluating the quality of voice input recognition by a voice portal and provides a novel and non-obvious method, system and apparatus for evaluating the quality of voice recognition by dynamic voice portals.

[0010] In a method of evaluating the quality of voice input recognition by a voice portal, a current grammar is extracted from the voice portal. A test input is generated for the current grammar. In this regard, the test input includes a test pattern and a set of active grammars for the current grammar. The test input can be entered into the voice server and the test pattern can be analyzed against the set of active grammars with a speech recognition engine in the voice server. Consequently, a measure of the quality of recognition for the current grammar can be derived.

[0011] Systems consistent with the present invention include a system for evaluating the quality of voice input recognition by a voice portal. An analysis interface extracts a set of current grammars from the voice portal. A test pattern generator generates a test input for each current grammar. The test input includes a test pattern and a set of active grammars corresponding to each current grammar. The system further includes a text-to-speech engine for entering each test pattern into the voice portal. A results collector analyzes each test pattern entered into the voice portal with the speech recognition engine against the set of active grammars corresponding to the current grammar for said test pattern. A results analyzer derives a set of statistics of a quality of recognition of each current grammar.

[0012] Additional aspects of the invention will be set forth in part in the description which follows, and in part will be obvious from the description, or may be learned by practice of the invention. The aspects of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The accompanying drawings, which are incorporated in and constitute part of the this specification, illustrate embodiments of the invention and together with the description, serve to explain the principles of the invention. The embodiments illustrated herein are presently preferred, it being understood, however, that the invention is not limited to the precise arrangements and instrumentalities shown, wherein:

[0014] Figure 1 is a block diagram illustrating an exemplary voice portal;

[0015] Figure 2 illustrates a voice portal with a system arranged in accordance with the principles of the present invention for evaluating the quality of voice input recognition by the voice portal; and

[0016] Figure 3 is a flowchart showing the process of evaluating the quality of voice input recognition by a voice portal.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0017] The present invention is a method and system for evaluating the quality of voice input recognition by a voice portal. The invention works by collecting a set of grammars for one or more voice applications disposed in a voice portal and testing the ability of the voice portal to recognize a particular grammar from among the set of other grammars that may be active with the particular grammar being tested. A measure of quality of recognition can be derived for each grammar, thereby enabling the voice portal to be reconfigured to allow for better voice input recognition.

[0018] Figure 2 illustrates a voice portal with a system arranged in accordance with the principles of the present invention for evaluating the quality of voice input recognition by the voice portal. The overall integrated system 100 can include a voice portal having a portal server 105 and a voice server 110. The portal server includes a voice aggregator 107 and one or more voice applications or portlets 108. The voice server can also include a text-to-speech (TTS) engine 114 and a signal manipulator 112. To this overall system 100, the present invention couples an analysis interface 120 to the portal server 105, a test pattern generator 125, a result collector servlet 130, a grammar and dependencies collector 140 and grammar database 145, a measurements results database 150, and a results analysis unit 152 which produces one or more reports 155.

[0019] The portal server 105 can be voice-enabled through coupling to a voice server 110. The voice server 110 is the unit with which an outside caller directly communicates, and can be linked to a telephone network or some other

communications network. The voice aggregator 107 is the software that manages the various voice applications 108 running on the portal server 105. When a user communicates with the voice portal, the voice aggregator presents the user with a menu, such as the menu in FIG. 1, wherein the user can select voice applications and content from a variety of selections, and can also navigate through the menu and the various voice applications 108. Each command that a user enters into the voice portal is a grammar which the voice server 110 must recognize to send an appropriate command to the portal server 105.

[0020] The analysis interface 120 exposes external entities to the logic of the voice aggregator 107 and allows the grammar and dependencies collector 140 to collect the various grammars enabled in the voice portal. The grammar database 145 and measurements results database 150 can be one or more data storage media or devices. The signal manipulator 112 can be any signal processing component that emulates the influence of different telephone or communications network qualities, such as line length, crosstalk, or noise, that is applied to the output of the TTS generator 114. The TTS 114 and manipulator 112 can be separated from the voice server 110 or can be integral to the voice server 110.

[0021] As used herein, a "current grammar" shall mean any grammar that is on the system of the voice portal, and can be any one of the grammars that corresponds to the various menu options for: (i) navigating through the voice portal, and (ii) selecting one of the portlets 108 on the portal server 105. The core idea of the invention is to check all current grammars in a voice portal with an automatic mechanism, so as to assess the capability and quality of voice recognition of the voice portal.

[0022] Figure 3 is a flowchart showing the process of evaluating the quality of voice input recognition by a voice portal. In this process, the present invention first provides for an analysis interface 120 software to be coupled to the voice portal and with the portal server 105. The analysis interface 120 communicates with the voice aggregator 107 to extract and retrieve any and all current grammars, at step 210. Since portal servers like portal server 105 are most likely implemented as a web application, the grammar and dependencies collector 140 could send one or multiple HTTP requests through analysis interface 120 to collect the current grammars, as well as the dependencies between the grammars. A database 145 can be used to store the data.

[0023] The test pattern generator 125 software can select a grammar from the set of current grammars stored in database 145, as well as the other grammars dependent on the selected grammar. A dependent grammar is any other grammar that may be executed by a user at any given aggregation step when navigating through the menu of the voice portal . Taking the menu in FIG. 1 as an example, if the user had navigated to menu item 17 for "Notes", the voice aggregator 107 could present a set of "active" grammars to a user at that stage, being the grammars for each of the portlets 30, 32, and 34, for "Projects", "Meetings" and "Directory", respectively, the grammars for the other menu options 16, 18, and 19, for "Information", "Directory" and "Sports", respectively, and the navigational grammars, such as "go back" or "quit." Therefore, as used herein, for any given current grammar which may be selected by the test pattern generator 125, the set of "active" grammars are all other grammars that may be presented to the user, including the selected current grammar, at the stage in the voice

portal where the user may enter a command corresponding to the selected current grammar.

[0024] For each selected current grammar, the test pattern generator 125 creates a "test input" for the grammar, at step 220. The test input can include both a test "pattern" and a set of active grammars corresponding to the current grammar for which the test input and test pattern is generated. The test pattern can be the actual word or term for the current grammar, or may also include additional words, terms, or sounds. The test pattern can also be entire sentences or phrases. Thus, the test input can include one or more test patterns that incorporate the selected current grammar in some way.

[0025] The test pattern generator 125 thus generates a test input for each current grammar and also aggregates a set of active grammars corresponding to the current grammar for each test input. The test input can be a VXML document having the test patterns and set of active grammars incorporated therein.

[0026] The test input is then entered into the voice server at step 230. The test pattern itself is entered through the TTS engine 114 and signal processor 112 into the voice server 110. The signal processor 112 can manipulate the sound of the test pattern by emulating the effects of different user voices, different languages, varying communications network qualities, and other modifications of the sound signature of the test pattern. Both TTS engine 114 and signal manipulator 112 may be separate units outside of the voice portal , in which case the synthesized output of the two units could be connected to the voice server 110 through some communications network. Or, the TTS 114 and signal manipulator 112 may already be integrated within the voice server

110. The set of active grammars corresponding to the current grammar for which the test pattern is generated is entered into the voice server 110 through a separate channel, such as from the results collector servlet 130, and may be done through the VXML test document described hereinabove.

[0027] Once the test pattern is entered into the voice server 110, in step 240, a speech recognition engine in the voice server can be used to obtain an assessment of how well the voice portal recognized the test pattern. The quality of the recognition of the test and the current grammar being tested by the test input is therefore obtained. This quality of recognition can be monitored and collected by the results collector servlet 130 and stored in the measurements results database 150. The quality of recognition can include a set of statistics that are generally used to assess the quality. Two examples of such statistics are the confidence level and n-best results, which generally used by speech recognition engines. Thus, the set of statistics can include a confidence level and a set of n-best results for the test input for each grammar tested, and resulting the confidence level and set of n-best results for the test input can be compared with an expected value for each metric to assess the quality of recognition.

[0028] In step 250, the process determines whether the quality of recognition is acceptable. If the quality is not acceptable, system 100 can be used to adjust and modify the selected current grammar, re-execute the test phase by running through steps 210, 220, 230 and 240, and re-assess whether the quality of recognition is acceptable. If the results are found to be acceptable at step 250, the process terminates.

[0029] An example of the process of the method of the present invention can be illustrated using the voice portal menu 10 of FIG. 1. To test the quality of recognition of menu item 34 for "Directory", the test input having a test pattern including the word "Directory" can be generated. When the current grammar for menu option 34 for "Directory" was extracted, the set of active grammars would also have been created. If the system on the voice portal is configured to have the grammars activated at all times for all directories in the Places 60 and Pages 65 levels of menu 10, as well as certain grammars for navigational commands like "Go back" and "Quit", the set of active grammars for the current grammar for portlet 34 for "Directory" would be: {"Business", "Entertainment", "Information", "Directory", "Sports", "Projects", "Meetings", "Directory", "Go Back", "Quit"}. A test pattern of "Directory" could be recognized by the speech recognition engine in the voice portal by assigning confidence levels to each grammar in the set of active grammars. A theoretical example of such confidence levels are listed below in Table 1.

Table 1

Grammar	Confidence Level
"Business"	0.21
"Entertainment"	0.10
"Information"	0.32
"Directory"	0.98
"Sports"	0.28
"Projects"	0.26
"Meetings"	0.35
"Directory"	0.99
"Go Back"	0.08
"Quit"	0.12

[0030] Confidence levels of close to one are regarded as a near perfect match, whereas confidence levels of near zero are regarded as not a match. If more than one grammar in the set of active grammars were to produce very high confidence levels, each above a certain pre-determined threshold, then the quality of recognition could be assessed as poor, since the system could incorrectly recognize one grammar for another. This can be seen in the example set above, where the two grammars for "Directory" each produce confidence levels that are far above any other of the grammars. The voice portal would therefore recognize one of the two grammars having the high confidence level. But it would not be able to distinguish between the two. Thus, the system would show that the quality of recognition is low in that the voice portal would not be able to easily distinguish between two grammars for two different commands. Hence, the user's ability to navigate through the portal would be compromised.

[0031] The present invention therefore provides a method and system for evaluating the quality of voice input recognition by a voice portal. The present invention can execute a test of the voice portal very quickly, at relatively low cost, and with far greater ease than a human system administrator of a voice portal could otherwise do. The present invention could test all grammars in a system, even if the grammars were spoken in different languages, and even if a voice portal system administrator does not know the languages. Furthermore, because of the ability of TTS engines to render different voices (male, female, fast, slow...), the present invention can utilize the TTS engine to test the voice portal with a much more robust input, than a human administrator can otherwise do. Also, because of a speech recognition engine's more

fine-grained ability to characterize the similarity of two sounds, while a human system administrator could only determine whether a voice portal simply worked or did not work, the present invention can measure how much one sound differs from another to produce a more detailed assessment of the quality of recognition by a voice portal.

[0032] The present invention can be realized in hardware, software, or a combination of hardware and software. An implementation of the method and system of the present invention can be realized in a centralized fashion in one computer system, or in a distributed fashion where different elements are spread across several interconnected computer systems. Any kind of computer system, or other apparatus adapted for carrying out the methods described herein, is suited to perform the functions described herein.

[0033] A typical combination of hardware and software could be a general purpose computer system with a computer program that, when being loaded and executed, controls the computer system such that it carries out the methods described herein. The present invention can also be embedded in a computer program product, which comprises all the features enabling the implementation of the methods described herein, and which, when loaded in a computer system is able to carry out these methods.

[0034] Computer program or application in the present context means any expression, in any language, code or notation, of a set of instructions intended to cause a system having an information processing capability to perform a particular function either directly or after either or both of the following a) conversion to another language,

code or notation; b) reproduction in a different material form. Significantly, this invention can be embodied in other specific forms without departing from the spirit or essential attributes thereof, and accordingly, reference should be had to the following claims, rather than to the foregoing specification, as indicating the scope of the invention.